VIRTUAL PRIVATE NETWORKS WITHIN A PACKET NETWORK HAVING A MESH TOPOLOGY

FIELD OF THE INVENTION

[0001] This invention relates to packet-switched networks and, in particular, to virtual private networks within packet networks having a mesh topology.

BACKGROUND OF THE INVENTION

[0002] The evolution of computer networking has seen a trend towards greater use of packet-switched networks having a mesh topology. To some extent, networks having a ring topology have begun to fall out of favour. Nevertheless, the ring topology provides certain attractive benefits, including resiliency and efficiency, that are not necessarily present in a mesh network.

[0003] In the interests of privacy and confidentiality it is sometimes desirable to establish an Ethernet virtual private network (VPN) over a packet-switched mesh network. A known approach for establishing Ethernet over a packet-switched mesh network includes tunneling each node to each other node, creating an N^2 mesh of tunnels. Other approaches include the "Martini" architecture or the RFC2547 architecture. These latter approaches are complex, hard to scale, and fail to adequately address resiliency and broadcast problems.

[0004] Accordingly, there remains a need for a method of establishing a flexible VPN in a mesh network that addresses some of the shortcomings of known solutions.

SUMMARY OF THE INVENTION

[0005] The present invention provides for the creation and management of a flexible virtual private network within a packet network having a mesh topology. The virtual private network may expand or contract dynamically by adding or dropping member nodes and dynamically re-determining its topology. The present invention employs label switched paths to create flexible virtual private networks within a mesh network.

[0006] In one aspect, the present invention provides a method of forming a virtual private network within a mesh network of nodes, the virtual private network includes member nodes selected from the network of nodes. The method includes the steps of distributing a membership message to the member nodes, the membership message including a VPN identifier; at each member node, determining a topology for the virtual private network, wherein for each of the member nodes the topology identifies at least one adjacent member node; and creating label switched paths between the member nodes and their adjacent member nodes, thereby establishing the virtual private network having the topology.

[0007] In a further aspect, the present invention provides a computer program product having a computer-readable medium tangibly embodying computer executable instructions for creating a virtual private network within a mesh network of nodes, the virtual private network including member nodes selected from the network of nodes. The computer executable instructions include computer executable instructions for distributing a membership message to the member nodes, the membership message including a VPN identifier; computer executable instructions for determining, at each member node, a topology for the virtual private network, wherein for each of the member nodes the topology identifies at least one adjacent member node; and computer executable instructions for creating label switched paths between the member nodes and their adjacent member nodes, thereby establishing the virtual private network having the topology.

[0008] In yet a further aspect, the present invention provides a system for forming a virtual private network within a mesh network of nodes, the virtual private network including member nodes selected from the network of nodes. The system includes means for distributing a membership message to the member nodes, the membership message including a VPN identifier; means for determining a topology for the virtual private network, wherein for each of the member nodes the topology identifies at least one adjacent member node; and means for creating label switched paths between the member nodes and their adjacent member nodes, thereby establishing the virtual private network.

[0009] In another aspect, the present invention provides a system for forming a virtual private network within a mesh network of nodes. The system includes member nodes selected from the network of nodes, wherein the member nodes receive a membership message, the membership message including a VPN identifier, and wherein the member nodes include a topology module for determining a topology for the virtual private network, wherein for each of the member nodes the topology identifies at least one adjacent member node; and label switched paths between the member nodes and their adjacent member nodes, wherein the label switched paths establish the virtual private network.

[0010] Other aspects and features of the present invention will become apparent to those ordinarily skilled in the art upon review of the following description of specific embodiments of the invention in conjunction with the accompanying figures.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] Reference will now be made, by way of example, to the accompanying drawings which show an embodiment of the present invention, and in which:

[0012] Figure 1 shows in diagrammatic form a system having a closed-loop label switched path established within a mesh network;

[0013] Figure 2 shows in diagrammatic form a virtual ring, according to the present invention;

[0014] Figure 3 shows the virtual ring of Figure 2 employed for point to point communication;

[0015] Figure 4 shows the virtual ring of Figure 2 employed for broadcast communication;

[0016] Figure 5 shows the virtual ring of Figure 2 employed for distributing control information;

[0017] Figure 6 shows, in flowchart form, a method of creating a virtual ring within a mesh network, according to the present invention;

[0018] Figure 7 shows, in flowchart form, a method of adding a new member node to a virtual ring;

[0019] Figure 8 shows, in diagrammatic form, a step in the method of adding a new member node to a virtual ring;

[0020] Figure 9 shows, in diagrammatic form, a further step in the method of adding a new member node to a virtual ring;

[0021] Figure 10 shows, in diagrammatic form, yet a further step in the method of adding a new member node to a virtual ring; and

[0022] Figure 11 shows, in diagrammatic form, another step in the method of adding a new member node to a virtual ring.

[0023] Similar reference numerals are used in different figures to denote similar components.

DESCRIPTION OF SPECIFIC EMBODIMENTS

[0024] The following detailed description of specific embodiments of the present invention does not limit the implementation of the invention to any particular communications protocol or language. Any limitations presented herein as a result of a particular type of communications protocol or language are not intended as limitations of the present invention.

[0025] The following detailed description includes specific embodiments of the present invention which establish a VPN having a ring topology. The present invention is not limited to ring-based VPNs. It will be understood that other VPN topologies may be realized, including tree-based topologies, such as in the case of a switched Ethernet LAN.

[0026] Reference is first made to Figure 1, which shows in diagrammatic form a system 10 that includes a mesh network 12 and a plurality of users 14 (shown individually as 14a, 14b,…, 14g). The mesh network 12 interconnects the users 14. The mesh network 12 includes a plurality of nodes 16 (shown individually as 16a, 16b,

…, 16h) and a plurality of physical links 18 (shown individually as 18a, 18b, …, 18n). The physical links 18 interconnect the nodes 16 with each other. Each of the users 14 is connected to a node 16 so as to be coupled to the mesh network 12.

[0027] The users 14 are entities capable of network communications, such as, but not limited to, computers, servers, other networks. The nodes 16 are devices that manage the exchange of communications over the physical links 18 of the mesh network 12. The nodes 16 are label-switched capable devices, and may include, but are not limited to, routers, switches, etc.

[0028] In one embodiment, the nodes 16 are Multi-Protocol Label Switching/Generalized Multi-Protocol Label Switching (MPLS/GMPLS) capable devices. The mesh network 12 supports MPLS/GMPLS transport and protocols. The MPLS/GMPLS technology forwards packets of data using labels attached to each packet, without requiring intermediate nodes to look at the content of each packet. In an MPLS/GMPLS network, the IP addresses within a packet are not examined, allowing MPLS/GMPLS to encapsulate data in order to provide for private data traffic. The present invention is not limited to embodiments realized using MPLS/GMPLS transport and protocols and may be realized using other label switched protocols, including ASTN, OUNI, PNNI and others, as will be understood by those of ordinary skill in the art.

[0029] In an MPLS/GMPLS system, label switched paths (LSPs) can be established by defining a transition in label values across a set of label switched routers (LSRs). To set up an LSP, the appropriate label mappings are distributed to the appropriate LSRs through a path set-up protocol. The LSRs each maintain a forwarding table populated with entries tying an incoming interface and label value to an outgoing interface and label value. A variety of signalling protocols exist for distributing labels and for other signalling, including Border Gateway Protocol (BPG), RSVP, and others.

[0030] Referring still to Figure 1, there is shown a closed-loop sequence of label switched paths 20. The closed-loop sequence of label switched paths 20 comprises a number of individual node 16 to node 16 LSPs established over the physical links 18b, 18f, 18k, 18l, and 18g.

[0031] The closed-loop sequence of label switched paths 20 establishes a virtual private network having a ring topology connecting users 14b, 14c, and 14d. Each of the nodes 16b, 16c, and 16d, in the closed-loop sequence of label switched paths 20 is a member node. The closed-loop sequence of label switched paths 20 passes through intermediate nodes 16d and 16f between member nodes 16b and 16d. The VPN with a ring topology may be referred to herein as a virtual ring.

[0032] The users 14b, 14c, and 14d, may use the virtual ring to communicate with other users on the ring. The ring provides certain ring-based advantages to the users 14b, 14c, and 14d, including resiliency and quality of service improvements and broadcast capabilities.

[0033] Reference is now made to Figure 2, which shows in diagrammatic form a virtual ring 30, according to the present invention. The virtual ring 30 includes four member nodes 16 (Fig. 1) and a closed-loop sequence of LSPs 32 interconnecting the four member nodes 16 in a closed loop. Each of the four member nodes 16 has a unique label identifier, namely #4, #6, #8, and #9. These labels are used to refer to a specific one of the four member nodes 16.

[0034] The member nodes 16 each maintain a forwarding table 36 populated by data identifying the other member nodes 16 and any information required to forward data to each other member node 16. For example, a forwarding table 36 may specify the "cost" associated with forwarding data to a particular member node 16 in each direction around the ring, *i.e.* a cost *x* for clockwise and a cost *y* for counterclockwise.

[0035] The closed-loop sequence of LSPs 32 passes through various intermediate nodes 34 between pairs of the four member nodes 16. Each intermediate node 34 shown in Figure 2 identifies a pair of labels corresponding to input and output labels. For example, traffic entering the closed-loop sequence of LSPs 32 at member node 16 label #4 and traveling clockwise towards member node 16 label #6 encounters a first intermediate node 34a where the label #22 is swapped for the label #5. At a second intermediate node 34b the label #5 is swapped for the label #17.

[0036] Routing on the closed-loop sequence of LSPs 32 employs a two level label stack. The first or top level of the stack is the tunnel label, *i.e.* the label for the hop being traversed on the closed-loop sequence of LSPs 32. The second level is the exit member node 16 label. In some cases, it is necessary to employ a three level label stack so as to differentiate between different networks that are interconnected via the closed-loop sequence of LSPs 32. In such a case, the third level is the network differentiator.

[0037] Reference is now made to Figure 3, which shows the virtual ring 30 of Figure 2 employed for point to point communication. The virtual ring 30 is shared between multiple VPNs (labels #1, #2, and #3). A packet of data may be sent from a particular ingress member node 16, such as member node 16 label #4, to a particular egress member node 16, such as member node 16 label #8 using the three level label stack. For example, a node-to-node communication 40 from label #4 to label #8 is accomplished using a label stack having the form: [#22, #8, #3][data]. In the label stack, the top label #22 identifies the label of the LSP segment or tunnel for the first hop of the closed-loop sequence of LSPs 32; the second label, #8, identifies member node 16 label #8 as the egress point; and the third label, #3, specifies that the packet relates to VPN number 3.

[0038] Reference is next made to Figure 4, which shows the virtual ring 30 of Figure 2 employed for broadcast communication. A packet of data is sent from a particular ingress member node 16, such as member node 16 label #4 to all member nodes 16, using the three level label stack. A broadcast communication 42 employs a label stack having the form: [#22, *, #3][data]. In the label stack, the second label, *, is a wildcard indicating that all member nodes 16 are egress points, meaning that every member node 16 on the ring (other than originating member node 16 label #4) receives a copy of the packet and also forwards it along the closed-loop sequence of LSPs 32 to the next member node 16.

[0039] Reference is now made to Figure 5, which shows the virtual ring 30 of Figure 2 employed for distributing control information. Instead of broadcasting data, as depicted in Figure 4, the closed-loop sequence of LSPs 32 may be used to send a control message 44. In this case, the label stack takes the form: [#22, #0][data]. The second level label of #0 indicates that the message is a control message. If the control message was directed at only one member node 16, such as label #8, then the label stack would take the form: [#22, #8, #0][data]. In this case the label #0 appears at the third level since the egress member node 16 label #8 is required at the second level.

[0040] Reference is now made to Figure 6, which shows, in flowchart form, a method 100 of creating a virtual ring within a mesh network, according to the present invention. The mesh network includes a plurality of MPLS/GMPLS capable nodes interconnected by a variety of physical links. Users are connected to some of the nodes.

[0041] The method begins in step 102 with the distribution of ring membership to those nodes that are to become members of the virtual ring. This may be done using a variety of MPLS signalling protocols, such as I-BGP. In one embodiment, each of the member nodes receives a control message containing data of, or similar to, the following form:

<RouterID>, <RingID>, <relativePosition>

where <RouterID> is the IP address of the member node, <RingID> is a variable length octet string identifying the virtual ring, and <relativePosition> is a floating point number. In one embodiment, <relativePostion> is an ordinal indicating where on the virtual ring the member node is positioned relative to other member nodes. The computation of the order of nodes on the ring is not scalable and the optimal order can change with time. Therefore, the administrator establishing the virtual ring provides an order by specifying an ordinal for each member node. The ordinal tells a member node that it will be adjacent to the next larger and next smaller number, using modulo arithmetic. The administrator chooses the ordinal, *i.e.* the relative position of nodes on the ring, but does not choose the routing. In other embodiments, <relativePosition> is some other sortable value, such as a number or a letter.

[0042] The distribution of ring membership need not include relative position information. If relative position information is not included in the membership message, then the nodes will not have pre-assigned neighbours and they will need to determine their neighbours using a suitable algorithm. For example, they could determine their neighbours on the basis of minimizing the cost associated with the LSPs, *i.e.* minimizing the circumference of the ring.

[0043] Following the distribution of ring membership, each member node selects a unique label for itself and attempts to identify the other member nodes on the virtual ring in step 104. The label may be selected randomly by each member node and a collision/de-clashing mechanism may be required to prevent any duplication of labels. Each member node seeks both the other member nodes' identities and their ordinals, if ordinals have been distributed. In one embodiment, each member node queries the BGP database for a list of other member nodes based upon the <RingID>. In response, each member node receives a list of the member nodes and their <relativePosition> ordinals.

[0044] In step 106, each member node computes a ring topology based upon the ordinals received in step 104. If ordinals have not been distributed, then each member node determines the ring topology based upon the applicable topology algorithm, *i.e.* minimizing cost of LSPs, and associated data gathered, *i.e.* regarding the cost of various routes between nodes. The ring topology tells the member node what the ring should look like from its perspective. In particular, the topology tells each member node which two member nodes are on either side of it.

[0045] In step 108, the member nodes initiate LSP set up with their adjacent member nodes to establish the closed-loop sequence of LSPs. In one embodiment, each member node sends an LSP set-up message to one of its adjacent member nodes; for example, the member node having the next highest ordinal, *i.e.* JOIN messages are sent clockwise around the ring. In another embodiment, each member node sends JOIN messages to the two member nodes on either side of it. To prevent the establishment of two LSP segments between a pair of nodes, the member nodes may send a JOIN message to an adjacent member node only if they have not yet received a corresponding JOIN message from that member node. In step 110, the member nodes receiving LSP set-up messages respond appropriately according to the signalling protocol employed in the set-up in order to establish the LSP segment between adjacent nodes. With an LSP segment established between each pair of adjacent member nodes on the virtual ring, a closed-loop sequence of LSPs emerges within the mesh network.

[0046] Once the closed-loop sequence of LSPs is established, the member nodes each build a forwarding table in step 112. To build the forwarding table the member nodes may, for example, send control messages around the ring gathering information about the other member nodes, including the "cost" associated with transmission across each LSP segment. The cost is a value that signifies the relative cost of using a particular LSP segment as compared to other LSP segments. It is the sum of the costs of the individual links that make up an LSP's cost and is based upon any number of factors, such as for example distance and bandwidth. A control message propagates around the ring gathering information until it returns to the sending member node where the information is extracted and used to populate the forwarding table at that member node.

[0047] Reference is now made to Figure 7, which shows, in flowchart form, a method 120 of adding a new member node to a virtual ring. The method 120 is described below in conjunction with Figures 8 to 11, which illustrate diagrammatically the progression of steps of the method 120 (Fig. 7) for the closed-loop sequence of label switched paths 20 of Figure 1. As will be seen in Figures 8 through 11, the closed-loop sequence of label switched paths 20 includes existing member nodes 16b, 16c, and 16d, and new member node 16a.

[0048] The method 120 begins in step 122 with the distribution of membership to the new member node 16a. As with the method 100 (Fig. 6) of creating the virtual ring, this may be done using a variety of MPLS/GMPLS signalling protocols, such as I-BGP. In one embodiment, the new member node 16a receives a control message containing data of, or similar to, to the following form:

<RouterID>, <RingID>, <relativePosition>

where <RouterID> is the IP address of the new member node, <RingID> is a variable length octet string identifying the virtual ring, and <relativePosition> is the floating point number that identifies the new member node's relative position on the ring. The new member node 16a may query a database for the virtual ring, such as a distributed BGP database, to determine the identities of the existing member nodes 16b, 16c, and 16d and their ordinals.

[0049] The existing member nodes 16b, 16c, and 16d are also notified that the new member node 16a is to be added to the virtual ring. This notification may take place by virtue of a regular update of ring membership, such as for example through updates to a distributed database of ring members, like with the I-BGP signalling protocol. Other methods of notifying existing member nodes 16b, 16c, and 16d may be used.

[0050] In step 124, the ring topology is calculated. The new member node 16a determines the ring topology based upon the ordinal values of each member node 16a, 16b, 16c, and 16d. It thereby determines which two existing member nodes are its adjacent nodes 16b and 16d. Similarly, the existing member nodes 16b, 16c, and 16d determine where the new member node 16a fits within the virtual ring.

[0051] In step 126, and as shown in Figure 8, the new member node 16a sends a set-up message 50 to its two adjacent member nodes 16b and 16d. In step 128, the two adjacent member nodes 16b and 16d send a response message 52 to the new member node 16a acknowledging the set-up request, as shown in Figure 9. For example, if the signalling protocol used is RSVP-TE, the new member node 16a may send a PATH message to each adjacent member node 16b and 16d which will respond with RESV messages. The new member node 16a will not be spliced in yet, but any data received over these new segments will be forwarded appropriately.

[0052] Once the new member node 16a has received both responses 52 from the adjacent member nodes 16b and 16d, then in step 130 LSP segments 54 and 56 between the new member node 16a and its two adjacent member nodes 16b and 16d, respectively, are established, as shown in Figure 10. If the signalling protocol used is RSVP-TE, this step may involve sending PATH refresh message with a special "splice" indication or subcode to the two adjacent member nodes 16b and 16d and the execution of the splice operation by those adjacent member nodes 16b and 16d to establish the LSPs 54 and 56 to the new member node 16a. The old LSP segment between the two adjacent member nodes 16b and 16d is still in place.

[0053] Once the new member node 16a has confirmed that the LSP segments 54 and 56 with its two adjacent member nodes 16b and 16d have successfully been established, it sends a clean-up message to them. In step 132, upon receipt of the clean-up message, the two adjacent member nodes 16b and 16d drop the old LSP segment 58 between them, as shown in Figure 11. At this point, the closed-loop sequence of label switched paths 20 has now been enlarged to splice in the new member node 16a. Steps 126 through 132 implement a "make-before-break" principle to minimize packet loss during ring contraction and expansion.

[0054] In step 134, the new member node 16a sends a control message around the ring gathering information regarding the identity of the other members and the costs associated with the LSPs between them. Once this information is received, it is used by the new member node 16a to populate its forwarding table and to select an appropriate unique label for itself. Then, in step 136, the new member node 16a sends another control message having the complete member information, including costs and labels, around the ring to allow other member nodes 16b, 16c, and 16d to update their own forwarding tables with the new information. In one embodiment, where the CR-LDP signalling protocol is used, the control messages are a QUERY-LABELS message and other hop by hop control messages, respectively.

[0055] It will be understood from the foregoing description that the make-before-break principle is also used in managing the removal of a member node from the virtual ring. For example, if a member node were to be removed from the ring, the two members adjacent the departing member node would recognize that they need to establish a direct LSP segment between them. Accordingly, the two adjacent members would set-up a new LSP segment and, once established, collapse the LSP segments with the departing member, thereby removing it from the virtual ring.

[0056] The virtual rings created within packet networks having mesh topology may be connected together to achieve a greater reach while maintaining a reasonable diameter. A possible application of the virtual rings includes connecting true resilient packet rings (RPR) over an MPLS/GMPLS wide area network (WAN). The virtual rings would thereby extend the resiliency and fairness characteristics of the RPRs into the WAN.

[0057] It will be understood by those of ordinary skill in the art that the virtual rings may be used in a hierarchical fashion. For example, a first rings may have a segment that traverses segments of a second ring. The rings are thus nested at the intersection and the third level label on the inner ring is actually used as the first level label on the outer ring.

[0058] Although the above description at times refers to particular signalling protocols, such as BGP, it will be understood that the present invention is not limited to a particular label switched signalling protocol.

[0059] It will also be understood that the present invention is not limited to ring topologies, but is applicable to tree-and-branch and other topologies of virtual private networks. Those of ordinary skill in the art will appreciate that with alternative topologies, like a tree-and-branch architecture, alternative methods are used to dynamically determine the topology of the VPN at each of the member nodes.

[0060] The present invention may be embodied in other specific forms without departing from the spirit or essential characteristics thereof. Certain adaptations and modifications of the invention will be obvious to those skilled in the art. Therefore, the above discussed embodiments are considered to be illustrative and not restrictive, the scope of the invention being indicated by the appended claims rather than the foregoing description, and all changes which come within the meaning and range of equivalency of the claims are therefore intended to be embraced therein.